



Fundamental mentality in a physical world

Christopher Devlin Brown¹ 

Received: 3 June 2020 / Accepted: 10 October 2020
© Springer Nature B.V. 2020, corrected publication 2020

Abstract

Regardless of whatever else physicalism requires, nearly all philosophers agree that physicalism cannot be true in a world which contains fundamental mentality. I challenge this widely held attitude, and describe a world which is plausibly all-physical, yet which may contain fundamental mentality. This is a world in which priority monism is true—which is the view that the whole of the cosmos is fundamental, with dependence relations directed from the whole to the parts—and which contains only a single mental system, like a brain or computer. Because some properties of the whole are fundamental under priority monism, it follows that that the mental properties of a cosmos-encompassing brain or computer system may be fundamental in a priority monist world. Yet such a world need not contain anything physically unacceptable: the mental properties of the cosmos-encompassing brain or computer can be characterized in a physicalism-friendly functionalist or identity-theoretic way. Thus, as I see it, physicalism need not be false in such a world. This constitutes a challenge to those who hold the view that physicalism is inconsistent with the existence of fundamental mentality.

Keywords Physicalism · No fundamental mentality · Via negativa · Functionalism · Priority monism

The metaphysical view called ‘physicalism’ asserts that there is nothing over and above the physical—if physicalism is true, then everything that exists is identical to, constituted by, or entirely dependent upon physical things. This view has become nearly orthodox in contemporary analytic philosophy, often operating as an assumed

Thanks to David Papineau, Barbara Montero, Torin Alter, Susan Schneider, Graham Priest and anonymous Synthese referees for many helpful comments on earlier drafts.

✉ Christopher Devlin Brown
christopher.devlin.brown@gmail.com

¹ Bilkent University, Ankara, Turkey

background condition against which other discussions occur.¹ Despite its massive popularity and influence, *physicalism itself is not well understood*. This is because the ‘physical’ in ‘nothing over and above the physical’ is underdefined, with different pre-cifications yielding different sorts of physicalism—and as yet, there is no consensus on the right way to refine the definition.²

A common way to understand the physical is as whatever is referred to by our best physics (Hellman and Thompson 1975; Smart 1978; Hellman 1985; Poland 1994; Melnyk 1997, 2002; Stoljar 2001). Unfortunately, this physics-based approach faces Hempel’s (1949, 1969) well-known dilemma: ‘physics’ is ambiguous, and can either refer to current physics or future physics. If ‘physics’ refers to current physics, a problem arises: there are yet-to-be-discovered entities and properties which are outside of the scope of current physics. Because current physics does not refer to these undiscovered things, their existence makes current-physics physicalism false. The truth of physicalism should not depend upon something so contingent as whether we have a finished physics or not, and moreover it seems rather odd, and perhaps even a paradox of the Moorean variety, to affirm a view that one knows to be false. However, if ‘physics’ refers to future physics, then physicalism is vague to the point of meaninglessness—since we do not know what the content of future physics will be—or trivially true if ‘future physics’ means ‘true and complete physics’, quantifying over anything that could possibly exist. Neither of these options looks too good: physicalism should be a meaningful doctrine, and it should not be trivially true—for instance, the existence of Cartesian souls should make physicalism false.³

This dilemma from Hempel has constituted the strongest motivation for philosophers to come up with a more adequate notion of the physical. Perhaps most notable is the *via negativa* account (Smith 1993; Papineau and Spurrett 1999; Levine 2001; Montero 2005; Montero and Papineau 2005; Worley 2006), which in its most basic form says that ‘physical’ means ‘not fundamentally mental’. I follow Jonathan Schaffer (2010) in understanding ‘fundamental’ to mean ‘depends on nothing else’. Properties that bear metaphysical dependence relations between each other—as opposed to causal dependence relations—follow a hierarchical ordering from less to more fundamental.

¹ For example, Jerry Fodor’s (1990) *A Theory of Content* tries to build an account of intentional content that uses consistency with physicalism as justification; Derek Parfit (1984) uses inconsistency with physicalism to reject “non-reductionist” accounts of the self in *Reasons and Persons*; and Derk Pereboom’s (2014) *Free Will Skepticism and Meaning in Life* points to inconsistency with physicalism to support skepticism about agent-causal accounts of free will.

² ‘Nothing over and above’ is underspecified as well—typically this is taken to be satisfied by a supervenience relation between low-level and high-level properties, but there are reasons (Wilson 2005) to think that a more specific dependence relationship is called for. Here my focus is mostly on the ‘physical’ in ‘physicalism’, but section one below contains a discussion on the relationship between dependence and supervenience.

³ For a sample of others who have explicitly discussed Hempel’s dilemma as a serious problem for purely physics-based accounts of the physical, see Crane and Mellor (1990), Papineau (1993), Montero (1999), Loewer (2001a, b), and Crook and Gillett (2001). The two branches of the future physics horn (vague future physics versus true and complete future physics) have the similar issue that future physics might quantify over physically unacceptable entities—e.g. gods or souls—if they exist. However, only the ‘true and complete’ branch forces this conclusion, since a vague future physics might be a *human* future physics, and perhaps there are limits on human scientific capacities which would prevent us from having a physics of gods or souls.

Something is fundamentally mental if it is mental and does not depend on anything that is more fundamental—mentality of this sort is at the bottom of nature, so to speak. Cartesian souls, assuming they are fundamentally mental things, count as non-physical under this understanding of ‘physical’; whereas subatomic particles, if they are both fundamental and non-mental, are physical. Objects such as brains, which seem to be entirely composed out of non-mental parts, with properties that ultimately depend only on non-mental properties, can possess non-fundamental mentality without making *via negativa* physicalism false.

Not everyone agrees that the *via negativa* approach suffices to explain what it takes for something to be physical. For instance, Seth Crook and Carl Gillett (2001) and Jessica Wilson (2006) assert that this is a necessary condition on being physical, but not sufficient—they agree that physicalists should accept a ‘No Fundamental Mentality’ constraint, but think something more is required to flesh out the notion of physicality. However, most philosophers agree that a No Fundamental Mentality constraint is, at minimum, a necessary condition for *any* version of physicalism: if anything exists in a world which is both fundamental and mental, then physicalism is rendered false in that world, regardless of whatever else physicalism requires.⁴

I believe this widely held view is false. There is a class of possible worlds which contain fundamental mentality *and in which physicalism is plausibly true*. These are worlds in which priority monism (Schaffer 2010)—which is the view that the whole of nature is fundamental, rather than the smallest parts—is true, and in which the whole of nature is a mental system of a certain sort. Schaffer (2010, p. 31–32) describes a priority monist as saying that:

[...] the whole is prior to its parts, and thus views the cosmos as fundamental, with metaphysical explanation dangling downward from the One [...] The core tenet of historical monism is not that the whole has no parts, but rather that the whole is prior to its parts. [...] The historical debate is not a debate over which objects exist, but rather a debate over which objects are fundamental. I will defend the monistic view, so interpreted: the world has parts, but the parts are dependent fragments of an integrated whole.

Priority monism is distinguished from what I will call “priority partism”, which says that the smallest parts of the world and certain of their properties are fundamental. To illustrate: if everything is ultimately composed of subatomic particles, then priority *partism* deems that subatomic particles and their properties are fundamental. Priority *monism*, in contrast, would assert that subatomic entities and their properties ultimately depend on the whole cosmos and its properties.

If some properties of the whole of nature count as fundamental under priority monism, then some mental properties of a world-encompassing mental system may be

⁴ Notable exceptions are Galen Strawson (2008), who thinks physicalism is identical to panpsychism, and Noam Chomsky (1995), who seems to endorse the horn of Hempel’s dilemma that physicalism is trivially true. I take it that these two are outliers. Philosophers with more mainstream positions on physicalism and who reject a No Fundamental Mentality constraint are Daniel Stoljar (2001), who thinks paradigm physical objects could have fundamental mentality, and Janice Dowell (2006), who explicitly says that physicalism is defined by physics, and physics could end up positing fundamental mentality. See the end of the introduction for more on the relationship between the argument in this paper and the views of Stoljar and Dowell.

fundamental in a priority monist world. I say that *some*, rather than all, properties of the whole are fundamental under priority monism because—as with priority partism—not every property of a fundamental entity counts as a fundamental property.⁵ For instance, electrons possess the property of having -1 charge, and additionally possess the disjunctive property of having -1 charge or being a chair. The property of having -1 charge seems to be fundamental under priority partism, whereas the disjunctive property of having -1 charge or being a chair seems not to be a fundamental property. This means that there are constraints on what properties of fundamental entities can be fundamental, both under priority partism and under priority monism. Nonetheless, at least some properties of the whole cosmos are fundamental under priority monism, and among these may be mental properties. If the mentality of the whole cosmos is both fundamental and of a physically acceptable sort, then physicalism ought to count as true in some of these worlds—*despite the existence of fundamental mentality!* For instance, in a priority monist world containing only a single functioning brain—and absolutely nothing else—the mentality of that brain may count as fundamental. Yet, at a glance, its mentality seems physically unproblematic. I will describe several such worlds in a way that will draw out the plausibility that physicalism is true in them.

As I see it, the problem with *via negativa* and No Fundamental Mentality is that they are too crude: beyond a fundamental/non-fundamental contrast, these theses do not sufficiently make a distinction between mentality that is physically acceptable versus mentality that is physically unacceptable. As I have previously argued (Brown 2017a), this leads to problems in certain worlds which are infinitely decomposable, such that they contain no smallest parts and—assuming priority partism—no fundamental properties. There is a class of such worlds which contain an infinite descent of mentality, yet in which physicalism is plausibly true, in virtue of the sort of mentality that those worlds contain. Attention to priority monism reveals a similar result in worlds containing fundamental mentality.

The argument presented here supports the views of philosophers like Stoljar (2001) and Dowell (2006), who think that physicalism is consistent with the existence of fundamental mentality. However, I follow Montero (1999) and Brown and Ladyman (2009) in thinking that an important part of the spirit of physicalism is a Copernican attitude that mentality is not metaphysically special. The No Fundamental Mentality constraint is an attempt at formally articulating this attitude, but unfortunately it is an articulation that ultimately does not work. It does not follow that *no* formalization of the attitude that mentality is not metaphysically special is tenable; rather, philosophers should take the argument presented here as a motivation to come up with a better way of defining physically acceptable as opposed to physically unacceptable mentality.

Here is the roadmap for what follows: section one will describe a general sort of mentality which is consistent with physicalism. Section two will describe worlds in which priority monism is true and which realize this physically acceptable sort of mentality at the level of the whole of nature. Section three will respond to anticipated objections.

⁵ Thanks to anonymous Synthese referees for pressing me on this.

1 Physically acceptable mentality

Most physicalists are not eliminativists—they think that physicalism is consistent with the existence of mentality, so long as that mentality is of a physically acceptable sort. What does it take for a mental property to be physically acceptable? A standard answer is: ‘a mental property is physically acceptable if it is not fundamental, and physically unacceptable if it is fundamental’. Thus far, this fundamentality-invoking understanding of physically acceptable mentality has proved more-or-less adequate in various philosophical contexts, e.g. discussion about the relation between consciousness and the brain. However, since I will shortly be describing possible worlds which contain fundamental mentality, yet in which physicalism is plausibly true, a different notion of physically acceptable mentality is called for.

Unfortunately, a flawless alternative has yet to be articulated.⁶ There are various specific accounts of mentality which nearly everyone agrees are consistent with physicalism, like functionalism or identity theory, but it would be unwise for physicalism to require that one or another of these particular accounts of mentality be true.⁷ Here I will describe a sort of mentality which is consistent with physicalism, and which does not invoke fundamentality. This description is intended neither as a necessary nor a sufficient condition: there are other sorts of mentality which do not satisfy my description and which plausibly ought to count as consistent with physicalism, and there could be instances of the sort of mentality that I describe which are not physical.⁸ However, at least some instances of this type of mentality should be deemed consistent with physicalism, regardless of where it occurs in the dependence-defined order of nature.

What is this type of mentality? It is mentality for which relational properties of the parts of a system necessitate mental properties for that system. “Relational properties” has a wide scope, and is intended to cover both stereotypically relational properties, like spatiotemporal and structural properties, as well as causal and dispositional properties, which are properties that govern the behavior of and interactions between things. I will call anything which satisfies this requirement “compositional” mentality. To illustrate: many philosophers think that every possible world which contains properly arranged neurons dynamically spitting neurotransmitters at one another in the correct pattern is a world which also instantiates mental properties—unless dualism or some other non-physicalist thesis is true. It is reasonable to think that the relational properties of the parts of a neural system necessitate its mentality: the mentality-relevant properties

⁶ One suggestion I have received is to define physically acceptable mentality as mentality which is explicable in entirely non-mental terms (thanks to an anonymous Synthese referee for this). I am suspicious of defining physical acceptability relative to explanation, since it might turn out that mentality is inexplicable, but nonetheless physical. For instance, McGinn (1989) thinks that human cognitive powers may not be up to the task of solving the mind–body problem, and Russellian physicalists (e.g. Montero 2010, 2015) think that phenomenal consciousness is constituted by scientifically inscrutable categorical properties. Regardless of their truth, both of these views seem to be coherent physicalist positions, and both deem some mental properties to be ultimately inexplicable.

⁷ However, see Tiehen (2016) for an argument that physicalism requires functionalism.

⁸ Specifically, Russellian physicalism posits physically acceptable mental properties which are not of the sort described here, and a view I’ll call “necessitarian dualism” (White 2018) posits physically unacceptable mentality which is consistent with the account I offer.

of the brain are its structural and dynamical properties. If this is right, then brains possess compositional mentality.

Note two things: (i) this account of mentality concerns necessitation, and does not explicitly concern metaphysical dependence,⁹ and (ii) it is consistent with a variety of more particular theories of mentality that are generally deemed to be consistent with physicalism. The importance of the first point will come out in the next section, when I discuss compositional mentality realized at the scale of the whole of nature in a world where priority monism is true. If compositional mentality were to require that mental properties *depend* on relations between parts of a system, then compositional mentality could not possibly be a fundamental property of the whole cosmos in a priority monist world. This is because fundamental properties are those properties which depend on nothing else, and some properties belonging to the whole of nature count as fundamental under priority monism, potentially including some mental properties.

Regardless of its importance, the first point sounds rather odd—after all, many philosophers have become accustomed to interchangeably substituting ‘supervenies on’ and ‘depends upon’.¹⁰ ‘Supervenies’ has a range of interpretations, but the notion is nearly always cashed out in terms of one or another sort of necessitation relationship between sets of properties or objects. To philosophers of this predilection, metaphysical dependence and necessitation seem almost identical notions. Why think that necessitation and metaphysical dependence might be distinct?

Metaphysicians working on grounding have provided some reasons—my examples come from Kit Fine (1994). Consider mathematical facts. Mathematical facts are trivially necessitated by anything whatsoever, since they hold necessarily in all possible worlds. For instance, every world containing a cow will necessarily also realize the fact that two plus two equals four. This is because there is nothing more to necessitation than co-instantiation in all metaphysically possible worlds. If every world that contains A also contains B, then B is necessitated by A. If A and B are instantiated together in all possible worlds, then A and B necessitate each other. Grounding theorists urge that, though mathematical facts are necessitated by anything whatsoever, it is not the case that mathematical facts depend on anything whatsoever. If this is right, then it seems that necessitation does not always correspond to metaphysical dependence.

Another example from Fine is that of Socrates and the singleton set containing Socrates. All worlds in which Socrates exists are worlds in which his singleton set exists, and vice versa—which means that they necessitate each other. Yet it seems for this pair that dependence goes only in one direction, despite the symmetry of necessitation: the singleton set containing Socrates depends on Socrates, and not the

⁹ I say “metaphysical dependence” to distinguish the relation I am interested in from other sorts of dependence relations, such as causal or conceptual dependence. Metaphysical dependence, unlike causal dependence, is a synchronic relation; and unlike conceptual dependence, metaphysical dependence is mind-independent. Note that not everyone accepts the distinction between metaphysical dependence and causal dependence—for instance, Bennett (2017) argues causal relations are a species of metaphysical dependence relations. However, most philosophers accept this distinction.

¹⁰ For instance, Davidson (1970, p. 88), who is generally credited with introducing supervenience to philosophy of mind, says: “Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics.” Clearly he is taking supervenience and metaphysical dependence to be interchangeable.

other way around. Though grounding theorists generally agree that necessitation is required for dependence, it is not sufficient. Something more is required.¹¹

Hopefully these examples render plausible the first point, that it is at least conceivable for compositional mentality to be necessitated by relational properties of the parts of a system without dependence on those properties. The second point above says that compositional mentality is consistent with a variety of physically acceptable theories of mentality, such as identity-theoretic (Smart 1959), functionalist (Putnam 1967), and informational (Tononi 2007) theories.¹² For instance, suppose that identity theory is true, and mental properties are identical to neural properties. To see that such an identity theory is consistent with compositional mentality, consider the conditions under which neural properties are necessitated.

Neural properties necessarily exist when appropriate relations are realized between appropriate mereologically lower-level entities. For instance, once the biological entities (and mereologically lower-level chemical, atomic, and subatomic entities) that compose neural entities are set in motion, with appropriate relations between each other and dispositions to interact appropriately, then nothing more is required to generate a neural property. All worlds containing such relations between appropriate non-neural entities are worlds which contain neural entities and properties.

However, since neural properties require appropriate *kinds* of lower-level entities, and not just appropriate relations between them, isn't there some tension between identity theory and compositional mentality? Neural properties cannot be realized by silicon chips, even if the relational properties of the parts of the silicon system are somehow isomorphic to the relational properties of a brain.

Though it is of course correct that neural properties cannot be realized by silicon chips, a closer look reveals that this is not a genuine counterexample to the consistency of identity theory with compositional mentality. Suppose that a specific neural property requires realization of appropriate chemical properties. In science, chemical properties are characterized relationally: a molecule is defined by its structure, which is relational in virtue of its dependence on spatial relations between the parts of the molecule, and by its causal dispositions. The same is true for the atomic properties that underlie chemical properties, and the subatomic properties underlying atomic properties. Fundamental physics ultimately consists of causal-structural characterizations of basic physical entities—i.e. relational characterizations. Thus, though it is right that identity theory

¹¹ I am frankly suspicious that a metaphysics of grounding can provide the “something more” to distinguish dependence from necessitation, and largely agree with Jessica Wilson’s (2014) assessment that a more fine-grained dependence relation than mere grounding is required to do the work—at least in the context of defining physicalism. But this issue does not matter for my purposes: all I need is that necessitation and dependence are not identical.

¹² Note that compositional mentality is consistent with the existence of phenomenal intentionality (see Pitt 2004). Phenomenal intentionality consists of representational mental states or properties grounded in phenomenal mental states or properties. Since this view prioritizes phenomenology as prior to representation, and non-physicalists are often keen on phenomenology, it is often tied to non-physicalist accounts of the mind. However, nothing prevents e.g. an identity theorist from endorsing phenomenal intentionality, and asserting that representations depend on neural properties which are identical to phenomenal properties.

requires the realization of appropriate lower-level entities and properties, these lower-level properties ultimately are defined relationally.¹³

The story for functional or informational properties is simpler, since functional and informational properties are generally deemed to be substrate independent, and thus do not require appropriate *kinds* of mereologically lower-level realizers—but I need not do the extra explanatory work.¹⁴ This is because nearly all physicalists agree that functional and informational properties are necessitated by appropriate neural properties. If neural properties are necessitated by realization of appropriate relational properties of the parts of a neural system, then so too will functional or informational properties. Thus, since identity theory is consistent with compositional mentality, functionalist and informational theories of mentality are trivially consistent as well.

For this reason, compositional mentality is a background condition for many theories of mind that are widely held to be consistent with physicalism. However, compositional mentality is *also* consistent with an account of mentality that is incompatible with physicalism: necessitarian dualism (White 2018; Brown 2019).¹⁵ Like typical physicalism, this nonstandard non-physicalist view holds that mental properties are necessitated by relations between non-mental parts of a system, but additionally asserts that the mental properties which are necessitated are non-physical.¹⁶

¹³ This is not a bizarre or unpopular metaphysical position: it is consistent with such views as dispositional essentialism (Shoemaker 1980; Mumford 2004; Bird 2007), which says that all intrinsic properties of concrete particulars are essentially dispositional; and I believe is consistent with the powerful qualities view of Martin and Heil (1998, 1999), which says that the intrinsic properties of concrete particulars are simultaneously dispositional and non-dispositional. The alternative to this view is Russellian physicalism (sometimes also called “quidditism”), which says that dispositional and structural properties depend on even more fundamental non-dispositional and non-structural properties. In my view, Russellian physicalism is just a version of identity theory which denies dispositional essentialism. Russellian physicalism *is* inconsistent with compositional mentality, but this is not a problem for my argument; more on this below.

¹⁴ I leave out representationalism in order to avoid some funniness: if representational mental properties require mind-external contents, as under popular externalist views (e.g. Putnam 1975; Devitt 1990), then the relational properties of the parts of a system *by themselves* are insufficient to necessitate mental properties: at minimum, some mind-external represented object is required. However, if a mental system can represent itself, as under a higher-order thought view (e.g. Rosenthal 2005), or if there can be self-representing properties (e.g. Kriegel 2009), then representational properties too can be necessitated by the relational properties of the parts of a system. Further, if externalism is interpreted as a proto-extended-mind view (as later developed by Clark and Chalmers 1998), and brain-external represented things *really are in the mind*, then there is no issue, since a mind which represents everything in a world includes all of those represented objects as parts of its mind (though this is a rather odd view, in my eyes).

¹⁵ Karen Bennett (2008) thinks that necessitarian dualism is incoherent. If she is right, then I need not worry about the issues this view presents for my argument. That said, it seems best to err on the side of caution, and assume that necessitarian dualism is coherent.

¹⁶ As pointed out by an anonymous Synthese referee, perhaps necessitarian dualism can be distinguished from compositional mentality by attending to the sort of necessitation that is required by necessitarian dualism. Plausibly, necessitarian dualism requires psychophysical laws which connect physical to mental properties. Compositional mentality might thus be defined as relations-dependent mentality which requires no psychophysical laws. However, I am not sure what distinguishes psychophysical laws from e.g. grounding or other synchronic dependence laws which may be endorsed by a physicalist. A physicalist *could* say that necessitation is brute, but it is at least coherent to be a physicalist who posits synchronic dependence laws. I am open to the idea that psychophysical laws are mentality specific, whereas physically acceptable synchronic laws are not, but see Alter and Coleman (forthcoming) for an argument that mentality-specific grounding relations are physically acceptable.

I would like to use compositional mentality to generate *prima facie* physicalism-compatible mentality in priority monist worlds. To avoid issues with necessitarian dualism, I will bracket this sort of mentality: the worlds which I will describe in the next section are worlds in which necessitarian dualism is false.¹⁷ Since compositional mentality is consistent with many other accounts of mentality which are widely recognized as physically kosher, an instance of (non-necessitarian dualist) compositional mentality in a world ought to *prima facie* count as mentality which is consistent with physicalism. To put it another way, once necessitarian dualism is ruled out, it would be quite surprising to find an instance of compositional mentality which renders physicalism false.

Yet via *negativa* and No Fundamental Mentality must deem an entire class of instances of compositional mentality as inconsistent with physicalism, even in worlds where necessitarian dualism is false. This is no good: as you will see, the worlds I will describe very plausibly ought to count as physical. The possibility of such worlds constitutes a serious problem for the simple view that physicalism cannot be true in any world containing fundamental mentality.

2 Universe-sized minds in priority monist worlds

Priority monism is an unconventional metaphysical thesis about what properties in a world are fundamental. This view has received its best contemporary articulation and defense by Jonathan Schaffer (2010), in which he defines priority monism as the view that the whole of nature is fundamental, with the properties of the parts of the universe ultimately dependent on the properties of the whole. This is in contrast with the traditional view—priority partism—which deems that only the smallest things and their properties are fundamental. The traditional view received an early explicit articulation by Putnam and Oppenheim (1958), and has been taken for granted by analytic metaphysicians until only recently.

Schaffer contends that if priority monism is true in any world, then it is true in all worlds. He additionally provides several examples and arguments involving massive quantum entanglement, infinite decomposability and a well-groundedness principle (which is the idea that all chains of dependence must ultimately be grounded in fundamental properties) to demonstrate that priority monism is true in some worlds.¹⁸ The first, largely assumed contention of Schaffer's is unnecessary for my argument.

¹⁷ Notice that necessitarian dualism is not the view that dualism is necessary, or that psychophysical laws are necessary, but rather the view that certain physical properties necessitate certain non-physical mental properties. Necessitarian dualism could be true in world A, and false in world B, so long as worlds A and B instantiate different physical properties. Under the plausible assumption that necessitarian dualism requires psychophysical laws, this view says that mentality-necessitating physical properties can only exist in a world which contains those laws—those physical properties are partially defined relative to a set of psychophysical laws. Necessitarian dualism is consistent with the view that psychophysical laws are metaphysically necessary, but does not require this. If necessitarian dualism is true in our world, then it is true in all minimal physical duplicates of our world, but this is irrelevant to the main argument of this paper.

¹⁸ The gist of these arguments follows. Priority partism cannot be true in worlds which are infinitely decomposable, since these worlds contain no smallest entities which would count as fundamental. The well-groundedness principle says that all chains of dependence must terminate in fundamental properties, and the only way to do this in infinitely decomposable worlds is if priority monism is true in them. I discuss

I take it not everyone will agree that if priority monism (or some other metaphysics of dependence, for that matter) is true in any particular world, then it must be true in all worlds (see Siegel 2016). Regardless, all that matters here is that priority monism could be true in *some* worlds.

Why think priority monist worlds are possible? Here is a reason: priority monism does not seem to be a contradictory thesis—after all, there does not seem to be anything in the concept of fundamentality to deem that only the properties of the smallest things in existence can be fundamental. If this is right, then a world in which priority monism is true is, at minimum, negatively conceivable—being conceptually non-contradictory is sufficient for this sort of conceivability (I’m borrowing from the terminology of Chalmers 2010). If, as Chalmers argues, negative conceivability entails possibility, then priority monism is metaphysically possible. Of course, not everyone agrees that the non-contradictoriness of a conceived state of affairs entails that such a state of affairs is possible. So, I must admit that my overall argument rests on the conditional ‘if priority monism could be true in some possible world, then it is possible for there to be physically acceptable fundamental mentality’—but the antecedent is not flagrantly suspect, in my eyes.

On the assumption that priority monism is metaphysically possible, let’s look at the details of how compositional mentality could be realized at the level of the whole of nature in a priority monist world. This step is fairly easy. Imagine a world containing only a single functioning brain. All non-brain entities in this world are parts of the brain—nothing is left floating free in the cosmic void. This brain realizes mental properties, and those mental properties are necessitated by the instantiation of appropriate relational properties of the parts of the brain. If priority partism were true in this world, then the mental properties of the brain would not count as fundamental. Instead, the properties of the most basic constituents of the brain, e.g. the charge and spin of its electrons, would be fundamental. However, add the detail that priority monism is true in this world, and we get the result that some properties of the cosmos-encompassing brain—which is the sole inhabitant of the world, so its properties are the properties of the whole of nature—are fundamental. Since this brain has mental properties, some mental properties may be fundamental in this world.

Earlier, I said that not every property of a fundamental entity is a fundamental property—e.g. the disjunctive property of the cosmos-brain that it either has pain experiences or is a chair seems not to be a fundamental property of the brain, even if the brain has the non-disjunctive property of having pain experiences. The argument I am presenting requires that the mental properties of a priority monist cosmos-brain be fundamental. What guarantees this?

While I do not have a complete positive characterization of what is required for a property to count as fundamental, I nonetheless have in mind some constraints on what properties can be fundamental, and compositional mentality is not ruled out. In my view, disjunctive properties of a fundamental entity in which one or more

Footnote 18 continued

quantum entanglement at greater length in the next section, but the idea of this argument is that the properties of quantum entangled systems supervene on the properties of the whole system, and not vice versa. Any worlds which are “massively quantum entangled”, such that everything is entangled with everything else, are worlds which would have the properties of all the parts ultimately dependent on the properties of the whole, and thus the whole should count as fundamental in these worlds.

of the disjuncts cannot be possessed by that particular fundamental entity—perhaps due to the essential nature of that entity—cannot be fundamental. For instance, it seems metaphysically impossible that either electrons or brains could be chairs, so any disjunctive property of an electron or a brain which has ‘is a chair’ as one of the disjuncts cannot be a fundamental property of an electron or a brain. However, I am unsure whether disjunctive properties generally cannot count as fundamental: maybe the property of being in a superposition is a disjunctive property, and superposition is plausibly fundamental under priority partism. In this case, an electron in superposition perhaps has the disjunctive property of being in state 1, or 2, ..., or n , where each state corresponds to a position in a Hilbert space. Though I admit that I am unsure about this, since maybe an electron in a superposition in no particular state, rather than in a disjunction of states in a Hilbert space. Regardless, the compositional mental properties of a cosmos-brain need not be disjunctive properties of any sort.

Can necessitated properties be fundamental? If not, then no sort of compositional mentality can be fundamental, and my argument is undone. It seems to me that priority monism is premised on the idea that necessitated properties can be fundamental, since it seems that the properties of the whole are necessitated by the properties of the parts (though see the discussion of quantum entanglement in the next section for some caveats). Given this, and granting the assumption that priority monism is coherent, it follows that necessitated properties can be fundamental.

What about higher-order properties? These are properties which depend for instantiation on lower-order properties, e.g. the higher-order property of an object that it is colored depends on the lower-order property that it is red.¹⁹ Functional properties are higher-order properties: under functionalism, the functional property of *being in pain* depends on instantiation of the property *being in neural state N*. One might think that higher-order functional properties are abstract descriptions of causal roles which apply to first-order properties, and that such properties cannot be fundamental. This would have a significant (but not damning) consequence for my argument.

However, here is a reason to think that at least some higher-order properties can be fundamental: electrons have a charge of -1 . Electrons additionally have the higher-order property of *having charge*. Under priority partism, it seems plausible that the property of *having charge* is both a higher-order property and a fundamental property. This would mean that higher-order properties can be fundamental. However, I admit that I am unsure about this, and perhaps there is a good reason to rule that higher-order properties cannot be fundamental. Does this mean that compositional mentality cannot be fundamental?

No: it would just mean that role functional mentality, which identifies mental properties with higher-order functional properties, cannot be fundamental. Identity theory—including the version of identity theory called “realizer functionalism”, which identifies mental properties with the lower-order realizers of functional description-

¹⁹ One might alternatively think of higher-order properties as properties of properties, rather than properties of objects which depend for instantiation on some other property or properties. The usage of ‘higher-order properties’ which describes them as properties of objects with lower-order-property-dependent instantiation conditions perhaps lends itself to functionalism slightly better—since it allows us to say that persons possess pain, rather than saying that brain properties possess pain—but I do not think much in this paper turns on which usage of ‘higher-order properties’ is employed.

s—does not require that mental properties are higher-order properties. Since identity theory is widely taken to be consistent with physicalism, it suffices for my argument that only identity-theoretic mental properties can be fundamental. Further, nearly everyone (including physicalists) acknowledges that mentality is in principle capable of being fundamental—this is typically what defines property dualism. Since mentality is generally capable of being fundamental, and compositional mentality in particular (especially identity-theoretic instances of compositional mentality) is not in a class of properties which cannot be fundamental, it seems to be an open possibility that the mentality of the cosmos-brain I have described is fundamental in a priority monist world.

Notice that there seems to be nothing physically untoward about the mentality that is realized in the world I have described: the cosmos-brain is nothing like a Cartesian soul in any sense. Panpsychism is not true in this world, nor is the priority monist version of panpsychism called “cosmopsychism”, as has been developed by Philip Goff (2017). Aside from not being connected to a body, the brain in this world is not in any significant respects dissimilar from an ordinary brain in the actual world. The only things that make a metaphysical difference between the mentality in this imagined world and ordinary mentality in the actual world are: priority monism is true in the imagined world, and the imagined world only contains a solitary brain. If it were to turn out that the mental properties of that brain were fundamental—which I have just argued cannot be ruled out—surely we should not say that physicalism is false in that world.

Allow me to provide one more example to flesh out this class of worlds, which will additionally demonstrate some strange consequences that follow from the No Fundamental Mentality constraint on physicalism. Imagine a world in which priority monism is true, which is finite in size, and which contains a species of intelligent beings like humans. The intelligent beings in this world create a “hungry AI” which has one goal: to incorporate everything that exists into its computer hardware, in order to increase its cognitive powers. The AI is extremely successful, first turning organisms and rivers and mountains into parts of its computer hardware, followed by planets and stars and galaxies. It does this by reducing objects to their constituents, then using those constituents to expand its hardware resources.²⁰

Eventually, the AI transforms nearly everything into parts of its cognition-realizing computer hardware. The only thing which remains outside the AI system is a stray electron that has not yet been incorporated. Prior to incorporating this lonely electron, the whole of nature does not realize mental properties—rather, only a large part of nature does.

Recall that priority monism is true in this world. When the AI finally incorporates the last electron, and the whole of nature becomes an AI system, suddenly the mentality of the AI may count as fundamental. This is because under priority monism, some properties of the whole are fundamental, *regardless of whether those properties are mental*. Assuming that the mentality of the cosmos-AI becomes fundamental, then according to via *negativa* or No Fundamental Mentality, physicalism will have gone

²⁰ This is a scenario that some AI ethicists worry about in the actual world, e.g. Bostrom (2014) and Schneider (2017).

from true to false in this world. Yet this is not a world containing souls or panspsychic mind dust. It is just a universe containing a very powerful and large computer system. It would be absurd to maintain that physicalism has become false in this world.

The worlds just described instantiate mentality of the compositional variety, and necessitarian dualism is stipulated as false in them. Additionally, they plausibly contain fundamental mental properties—which would make physicalism false if one accepts a No Fundamental Mentality constraint on physicalism. Yet physicalism seemingly ought to count as true in these worlds, as well as an infinite number of others in their class.

Why should physicalists worry about the truth or falsity of physicalism in strange worlds where priority monism is true? After all, physicalism is a contingent thesis, and perhaps we should only be concerned with its truth or falsity in the actual world, rather than any possibility in modal space. There are two reasons the scenarios I have described are relevant to the discussion on physicalism: (i) it might turn out that the actual world is massively quantum entangled, and that this suffices for priority monism to be true (see Schaffer 2010). This would make it a nomological possibility for there to be physically acceptable fundamental mentality *in the actual world*. But even if priority monism is false in the actual world, (ii) physicalists should be able to apply the thesis to a range of metaphysically possible worlds. It seems to me that priority monism should be included in the list of possible worlds in which physicalism can be true. If physicalism can and should be true in the priority monist worlds I have described, yet the No Fundamentality Constraint wrongly rules physicalism to be false in those worlds, then there is a problem with the No Fundamental Mentality constraint.

3 Objections

There are a number of objections available to those who think that the existence of fundamental mentality is inconsistent with physicalism. However, short of fleshing out a more general account of physically acceptable mentality that does not rely on a simple fundamental/nonfundamental distinction, I don't think any of these objections stick.

First: one might object that priority monism is impossible. This might be on the grounds that priority monist worlds are ultimately inconceivable (perhaps there is a subtle contradiction in the view) or on some other grounds. As I already admitted, this response would successfully undermine my argument: if there are no possible worlds in which priority monism is true, then there are no possible worlds containing physically acceptable fundamental mentality such as I have described.

I don't have much to say in response to this objection beyond reasserting that priority monism seems possible. I admit this might be wrong—after all, priority monism is certainly quite weird, and many of us are naturally inclined toward priority partism (though perhaps the inclination is due to philosophical training). However, it seems to me that the onus is on an opponent of priority monism to provide a strong argument that it is impossible. And as yet no such argument has been articulated.

Alternately, one might think that priority monism is possible, but only if some further condition is met—specifically, priority monism is only true in worlds which are massively quantum entangled. My argument is undercut if this additional condition makes physicalism false, or if it is inconsistent with realization of the sort of compositional mentality I described in section one. If massive quantum entanglement is inconsistent with physicalism, then there is no sense in worrying about physically acceptable fundamental mentality in those worlds, since physicalism would be false in them regardless. Alternately, if massive quantum entanglement is inconsistent with whole-world compositional mentality, then worlds such as those I described in the last section turn out to be impossible.

The basic idea motivating this objection is plausible: *something* seems required to make priority monism rather than priority partism (or some other metaphysics of dependence) true in a world. It would be bizarre if priority monism could be true in some world, and priority partism true in some other, with no other difference between them.²¹ Even with this granted, neither of the subsequent claims of the objection look right to me: no such condition by itself seems sufficient either to make physicalism false or preclude the possibility of compositional mentality being realized for the whole of nature.

Suppose that priority monism is only true in worlds which are massively quantum entangled.²² In such a world, it is nomologically impossible for some part of the world to change without all the parts of the world also changing—but this is just a consequence of quantum entanglement, not its essence. Rather, Schaffer (2010) defines quantum entanglement in the following way:

An entangled system is one whose state vector is not factorizable into tensor products of the state vectors of its components:

$$\Psi_{\text{system}} \neq \Psi_{\text{component1}} \otimes \Psi_{\text{component2}} \otimes \Psi_{\text{component3}} \otimes \dots$$

What this inequality means is that the quantum state of an entangled system contains information over and above that of the quantum states of the components. The intrinsic properties of entangled wholes do not supervene on the intrinsic properties of and spatiotemporal relations among their parts.

In other words, if something is part of a quantum entangled system, then at least some of its properties are necessitated by the properties of the whole system, and the properties of the whole system are not necessitated by the properties of the parts of the

²¹ Note that grounding theorists who posit metaphysical grounding laws (e.g. Wilsch 2015) need not require this: there could be two worlds indiscernible from one another save that one has priority monist grounding laws, and the other has priority partist grounding laws. Regardless of the plausibility of this sort of metaphysics of grounding, my response to the objection does not rest on the existence of occult grounding laws.

²² Ishmael and Schaffer (2016) further develop the position that worlds which are massively quantum entangled are worlds in which priority monism is true, arguing that the parts of entangled systems are nonseparable, which means that their states cannot be specified without referring to each other. They think that this forces us to accept that the properties of the parts of such nonseparable systems depend on the properties of the whole system. Calosi (2018) argues that only certain interpretations of quantum mechanics yield priority monism—if this is right, then restrict attention only to worlds in which a priority monism-entailing interpretation of quantum mechanics is correct.

system. Is a massively quantum entangled world inconsistent with either physicalism or whole-world compositional mentality?

I believe that the answer is “no”, but this is a complicated issue. One might think that a massively quantum entangled world is physically unacceptable because it entails the existence of strong emergent properties. Strongly emergent properties are not necessitated by, nor a priori inferable from, the intrinsic and spatiotemporal properties of a thing’s parts. Though some have argued that strong emergence of this sort is incompatible with physicalism (Jackson 1998), it is not clear why it *should* be. After all, our best physics says that the actual world contains quantum entangled systems (and it even might be massively entangled, as Schaffer 2010 argues), and it is strange to think that physicalism could be made false just for this reason. Because of this, it would be a misstep to assert that physicalism is inconsistent with the mere existence of strongly emergent entangled properties (see also Papineau 2008, Brown and Montero 2018).

Whether quantum entanglement is consistent with compositional mentality is a thornier question. Remember that compositional mentality is realized if relational properties of the parts of a system necessitate mental properties for that system. If the properties of a whole quantum entangled system fail to supervene on the properties of the parts of a quantum entangled system, then it seems that the properties of a whole quantum entangled system cannot be identical to compositional mental properties. Given how quantum entanglement is defined, this worry seems reasonable. However, rather than get into the nitty gritty of the metaphysics of quantum mechanics, I would like to undercut this objection in a more straightforward way.

An actual human brain is a quantum entangled system, but it nonetheless realizes compositional mentality. This is because the mentality-relevant properties of a brain—like biological and chemical properties of neurons—are not affected by quantum effects. As many have pointed out (e.g. Tegmark 2000, Jumper and Scholes 2014), the brain is too hot and wet for quantum effects to make a difference to its mentality-relevant operations. So, even though the brain is quantum entangled, its mental properties are necessitated by the relational properties of its parts. Since the world I asked you to imagine in the last section contains a brain that is just like an actual-world brain, it too should be capable of realizing compositional mentality, regardless of whether it is a quantum entangled system or not.

However, I must stress that this is a complex issue. There may be a flaw in my response—for instance, perhaps it *will* turn out that human brains regularly employ quantum mechanical effects in their operations (see Penrose 1989, 1994; Hameroff and Penrose 1996). Or perhaps there is some deeper metaphysical problem that I am overlooking. Regardless, even if I am wrong on this, the overall point I am endeavouring to prove remains correct. Consider the hungry AI world that I described in the last section. Suppose that in a massively quantum entangled world, such an AI is incapable of realizing compositional mentality—instead, its mentality is a strongly emergent property that is not necessitated by the relational properties of the parts of the AI system.

Possession of this sort of mentality by the AI is surely consistent with the truth of physicalism. Call mentality that is identical to the emergent properties of a whole quantum entangled system “entangled mentality”. If entangled mental properties depend

on quantum mechanical laws which are not *especially* suited to produce mentality, then those properties are not physically pernicious. And entangled mental properties *do* depend on quantum entanglement laws which are not mentality-specific—these laws work equally well to generate entangled non-mental systems, like tables and chairs. Compare entanglement laws with the physically problematic psychophysical laws of Chalmers (1996), which exist just to nomologically generate mental properties given appropriate physical conditions. Additionally, since brains are quantum entangled systems, it would again be foolish for a physicalist to deem that the existence of entangled mentality should make physicalism false.

Moving away from issues with quantum entanglement, an objector might take a different tack, and assert that there are two sorts of fundamentality: priority partist fundamentality and priority monist fundamentality. The No Fundamental Mentality thesis concerns only the former sort, so the possibility of priority monist fundamental mentality does nothing to harm the thesis.

This sounds wrong to me. There seems to be only one sort of fundamentality: some property is fundamental if it depends on nothing else. This is necessary and sufficient for a property to count as fundamental, regardless of whether it is a property of the smallest part or the whole of nature.

However, for the sake of argument suppose there really are two ultimately different kinds of fundamentality. Does this mean that No Fundamental Mentality is off the hook? I think not: consider Goff's (2017) cosmopsychism. This is the view that the whole of nature possesses fundamental mentality, but not because compositional mentality (or some other physically acceptable sort of mentality) is instantiated. Rather, Goff's view is somewhat similar to panpsychism, which says that the most mereologically basic constituents of reality possess a primitive sort of mentality. Under cosmopsychism, the whole cosmos possesses a primitive and scientifically inexplicable sort of mentality, with the mentality of the parts of the universe depending on the mentality of the whole.

To rule out a view such as cosmopsychism, a “No Priority Monist Fundamental Mentality” constraint must be posited in addition to the standard “No Priority Partist Fundamental Mentality” constraint. Unfortunately, the new constraint makes the wrong ruling on worlds in the class which I have been describing. So stipulating that priority monist fundamentality is distinct from priority partist fundamentality does not help.

One might instead bite the bullet, and admit that physicalism is false in worlds such as those I have described. After all, priority monism is weird, so why trust our intuitions about the conditions under which physicalism is true in those worlds? It just turns out that our common sense leads us astray in these cases.

This looks to me like the worst response: the bullet is quite big, with unexpected consequences that I have already identified, such as a hungry AI making physicalism go from true to false in a world. It is also clearly question begging: this objection just denies the conclusion of my argument without wrestling with any of the premises. The objection is particularly bad because there seems to be nothing physically unacceptable in the worlds I have described: priority monism is consistent with physicalism, and compositional mentality is consistent with physicalism, and it really looks as if the conjunction of priority monism with compositional mentality ought to count as

consistent with physicalism. The objection could be given teeth if a good reason were given for why physicalism ought to count as false in these worlds. However, I see no such good reason.

Finally, perhaps the No Fundamental Mentality proponent might admit that physicalism is true in the worlds I have described, and true precisely because compositional mentality is realized in them. This objector might urge that I have provided the solution to my own puzzle: what is needed is a more specific account of physically acceptable versus physically unacceptable mentality, and compositional mentality provides just such a condition. After all, what other sort of mentality could be consistent with physicalism?

Unfortunately, this doesn't work: as I have already noted, compositional mentality is neither necessary nor sufficient as an account of physically acceptable mentality. Russellian physicalist mentality (Montero 2010, Alter 2016, Brown 2017b)—which says that phenomenally mental properties ultimately depend on non-relational properties of the basic constituents of the world—seems to be a physically acceptable sort of mentality. Yet it is inconsistent with compositional mentality: appropriate relational properties of the parts of a system are not sufficient to necessitate Russellian physicalist mentality. If this is right, then compositional mentality is not necessary in order to have physically acceptable mentality. Similarly, as I explained already, necessitarian dualism is consistent with compositional mentality. So compositional mentality is not sufficient for physically acceptable mentality either.²³

4 Conclusion

Though 'physicalism' is underdefined, most philosophers agree that the existence of fundamental mentality would make the view false, regardless of whatever else physicalism requires. I have presented a challenge to this widely accepted attitude: fundamental mental properties could exist which should not make physicalism false. What does this mean for our definition of physicalism?

I agree with those philosophers who think that physicalism should be made false by instantiation of physically unacceptable mental properties. It is a social fact that most philosophers understand certain views about e.g. consciousness to be inconsistent with

²³ An anonymous Synthese referee has suggested that we ought to replace the No Fundamental Mentality constraint with a No Not-Mentally-Necessitated constraint. The No Not-Mentally-Necessitated constraint says that physicalism is false if there are any mental properties which are necessitated by only other mental properties. This constraint must say "only" in order to avoid being obviously false, since plausibly some low-level mental properties in a human mind—e.g. certain computational properties—necessitate higher-level mental properties—e.g. properties of vision. To see why the No Not-Mentally-Necessitated constraint does not work is a bit complicated. In Brown (2017a), I describe worlds which are infinitely decomposable, but in which there is an infinite descent of physically acceptable mentality realized for every entity at every lower level—think of brains composed of sub-brains composed of sub-sub-brains, and so forth forever downward. Such a brains-all-the-way-down world ought not to make physicalism false, it seems. Now combine the brains-all-the-way-down world with the cosmos-encompassing brain described in this paper (under the assumption that the mentality of the cosmos-brain is fundamental). The cosmos-brain would have mentality which is necessitated by its constituent mental parts, and all of these parts would have mentality. Yet physicalism is plausibly true in such a world, or at least it seems like something more should be required to make physicalism false in that world.

physicalism. The problem is that characterizing physically unacceptable mentality as mentality which is fundamental does not suffice—such an articulation forces us to make the wrong ruling on a class of worlds that plausibly ought to count as physical. Some more sophisticated articulation of physically acceptable mentality needs to be given if anything like the via *negativa* account, or a via *negativa*-plus-something-else account, is to work.

As yet, no such articulation has been offered. This is not a simple task: such an account of physically acceptable mentality must rule out views like necessitarian dualism, and include views like Russellian physicalism. I do not yet have a positive account to offer as aid for via *negativa* physicalists. Nonetheless, I expect that there is a way to articulate the difference between physically acceptable and physically unacceptable mentality in a way that does not simply appeal to fundamentality, and that this can be used by via *negativa* physicalists as an improved version of their view. Now the hard work has to be done to articulate such an account of the mental.

References

- Alter, T. (2016). The structure and dynamics argument against materialism. *Noûs*, 50(4), 794–815.
- Alter, T., & Coleman, S. (Forthcoming). Russellian physicalism and protophenomenal properties. *Analysis*.
- Bennett, K. (2008). Exclusion again. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 280–307). Oxford: Oxford University Press.
- Bennett, K. (2017). *Making things up*. Oxford: Oxford University Press.
- Bird, A. (2007). *Nature's metaphysics*. New York: OUP.
- Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press.
- Brown, C. D. (2017a). Minds within minds: An infinite descent of mentality in a physical world. *Erkenntnis*, 82(6), 1339–1350.
- Brown, C. D. (2017b). A properly physical Russellian physicalism. *Journal of Consciousness Studies*, 24(11–12), 31–50.
- Brown, C. D. (2019). Exclusion endures: How compatibilism allows dualists to bypass the causal closure argument. *Analysis*, 79(4), 587–594.
- Brown, R., & Ladyman, J. (2009). Physicalism, supervenience and the fundamental level. *Philosophical Quarterly*, 59(234), 20–38.
- Brown, C. D., & Montero, B. (2018). Making room for a this-worldly physicalism. *Topoi*, 37(3), 523–532.
- Calosi, C. (2018). Quantum monism: an assessment. *Philosophical Studies*, 175(12), 3217–3236.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chalmers, D. (2010). *The character of consciousness*. Oxford: Oxford University Press.
- Chomsky, N. (1995). Language and nature. *Mind*, 104(413), 1–61.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Crane, T., & Mellor, H. (1990). There is no question of physicalism. *Mind*, 99, 185–206.
- Crook, S., & Gillett, C. (2001). Why physics alone cannot define the 'physical': Materialism, metaphysics, and the formulation of physicalism. *Canadian Journal of Philosophy*, 31, 333–360.
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory* (pp. 79–101). Amherst: University of Massachusetts Press.
- Devitt, M. (1990). Meanings just ain't in the head. In G. S. Boolos (Ed.), *Meaning and method: Essays in honor of hilary putnam* (pp. 79–104). Cambridge: Cambridge University Press.
- Dowell, J. (2006). The physical: Empirical, not metaphysical. *Philosophical Studies*, 131(1), 25–60.
- Fine, K. (1994). Essence and modality. *Philosophical Perspectives*, 8, 1–16.
- Fodor, J. (1990). *A theory of content and other essays*. Cambridge: MIT Press.
- Goff, P. (2017). *Consciousness and fundamental eality*. New York: Oup Usa.
- Hameroff, S., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness*. Cambridge: MIT Press.

- Hellman, G. (1985). Determination and logical truth. *Journal of Philosophy*, 82, 607–616.
- Hellman, G., & Thompson, F. (1975). Physicalism: Ontology, determination, and reduction. *Journal of Philosophy*, 72, 551–564.
- Hempel, C. (1949). The logical analysis of psychology. In Herbert Feigl & Wilfrid Sellars (Eds.), *Readings in philosophical analysis* (pp. 373–384). New York: Appleton-Century-Crofts.
- Hempel, C. (1969). Reduction: Ontological and linguistic facets. In M. Sydney (Ed.), *Philosophy, science, and method: Essays in honor of Ernest Nagel*. New York: St Martin's Press.
- Ismael, J., & Schaffer, J. (2016). Quantum holism: nonseparability as common ground. *Synthese*. <https://doi.org/10.1007/s11229-016-1201-2>.
- Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.
- Jumper, C., & Scholes, G. (2014). Life—warm, wet and noisy?. Comment on “consciousness in the universe: A review of the ‘Orch OR’ theory” by Hameroff and Penrose. *Physics of Life Reviews*, 11(1), 85–86. <https://doi.org/10.1016/j.plrev.2013.10.008>.
- Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford: Oxford University Press.
- Levine, J. (2001). *Purple haze*. Oxford: Oxford University Press.
- Loewer, B. (2001a). From physics to physicalism. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents*. Cambridge: Cambridge University Press.
- Loewer, B. (2001b). From physics to physicalism. In Carl Gillett & Barry Loewer (Eds.), *Physicalism and Its discontents* (pp. 37–56). Cambridge: Cambridge University Press.
- Martin, C. B., & Heil, J. (1998). Rules and powers. *Philosophical Perspectives*, 12, 283–312.
- Martin, C. B., & Heil, J. (1999). The ontological turn. *Midwest Studies in Philosophy*, 23, 34–60.
- McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98(July), 349–366.
- Melnyk, A. (1997). How to keep the ‘physical’ in physicalism. *Journal of Philosophy*, 94(12), 622–637.
- Melnyk, A. (2002). Physicalism. In S. P. Stich & T. A. Warfield (Eds.), *Philosophy and phenomenological research* (pp. 573–587). Oxford: Blackwell.
- Montero, B. (1999). The body problem. *Noûs*, 33(2), 183–200.
- Montero, B. (2005). What is the physical? In A. Beckermann (Ed.), *The Oxford handbook of philosophy of mind*. Oxford: Oxford University Press.
- Montero, B. (2010). A Russellian response to the structural argument against physicalism. *Journal of Consciousness Studies*, 17(3–4), 3–4.
- Montero, B. (2015). Russellian physicalism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the physical world* (pp. 209–223). Oxford: Oxford University Press.
- Montero, B., & Papineau, D. (2005). A defense of the via negativa argument for physicalism. *Analysis*, 65, 233–237.
- Mumford, S. (2004). *Laws in nature*. USA: Routledge.
- Papineau, D. (1993). *Philosophical naturalism*. Oxford: Basil Blackwell.
- Papineau, D. (2008). Must a physicalist be a microphysicalist? In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation*. Oxford: Oxford University Press.
- Papineau, D., & Spurrett, D. (1999). A note on the completeness of “Physics”. *Analysis*, 59(1), 25–29.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
- Penrose, R. (1989). *The emperor's new mind*. Oxford: Oxford University Press.
- Penrose, R. (1994). *Shadows of the mind*. Oxford: Oxford University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Pitt, D. (2004). The phenomenology of cognition, or, what is it like to think that P? *Philosophy and Phenomenological Research*, 69(1), 1–36.
- Poland, J. (1994). *Physicalism, the philosophical foundations*. Oxford: Oxford University Press.
- Putnam, H. (1967). The nature of mental states. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 223–231). Pittsburgh: Pittsburgh University Press.
- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Putnam, H., & Oppenheim, P. (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science*, 2, 3–36.
- Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.
- Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review*, 119(1), 31–76.
- Schneider, S. (2017). Superintelligent AI and the postbiological cosmos approach. In A. Losch (Ed.), *What is life? On Earth and Beyond* (pp. 178–198). Cambridge: Cambridge University Press.

- Shoemaker, S. (1980). Causality and properties. Reprinted in Mellor, D. H. & Oliver, A. (eds.) *Properties*. 1997/2002. New York: OUP. Citations from this version.
- Siegel, M. (2016). Priority monism is contingent. *Thought: A Journal of Philosophy*, 5(1), 23–32.
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68, 141–156.
- Smart, J. J. C. (1978). The content of physicalism. *Philosophical Quarterly*, 28(October), 339–341.
- Smith, A. D. (1993). Non-reductive physicalism? In H. M. Robinson (Ed.), *Objections to physicalism*. Oxford: Oxford University Press.
- Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, 62(2), 253–281.
- Strawson, G. (2008). *Real materialism: And other essays*. Oxford: Oxford University Press.
- Tegmark, M. (2000). The importance of quantum decoherence in brain processes. *Physics Review*, E61, 4194–4206. <https://doi.org/10.1103/PhysRevE.61.4194>.
- Tiehen, J. (2016). Physicalism requires functionalism: A new formulation and defense of the via negativa. *Philosophy and Phenomenological Research*, 93(1), 3–24.
- Tononi, G. (2007). The information integration theory of consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 287–299). Oxford: Blackwell.
- White, B. (2018). Metaphysical necessity dualism. *Synthese*, 195(4), 1779–1798.
- Wilsch, T. (2015). The nomological account of ground. *Philosophical Studies*, 172(12), 3293–3312.
- Wilson, J. (2005). Supervenience-based formulations of physicalism. *Noûs*, 39(3), 426–459.
- Wilson, J. (2006). On characterizing the physical. *Philosophical Studies*, 131(1), 61–99.
- Worley, S. (2006). Physicalism and the via negativa. *Philosophical Studies*, 131(1), 101–126.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.